UNCLASSIFIED

AD **431018**

DEFENSE DOCUMENTATION CENTER

FOR

SCIENTIFIC AND TECHNICAL INFORMATION
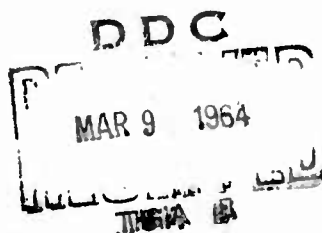
CAMERON STATION, ALEXANDRIA, VIRGINIA

UNCLASSIFIED

64-9

D1-82-0326

BOEING SCIENTIFIC RESEARCH LABORATORIES

Norms and Condition Numbers

Albert W. Marshall

Ingram Olkin

Mathematics Research

February 1964

NORMS AND CONDITION NUMBERS

by

Albert W. Marshall
Mathematics Research Laboratory

and

Ingram Olkin
Stanford University

SUMMARY

The condition number c_φ of a non-singular matrix A is defined by $c_\varphi(A) = \varphi(A)\varphi(A^{-1})$, where ordinary φ is a norm. It is known that for certain norms, the matrix AA* is more "ill-conditioned" than A, i.e., $c_\varphi(A) \leq c_\varphi(AA^*)$. We prove that this inequality holds whenever the norm φ is unitarily invariant ($\varphi(A)$ is a function of the characteristic roots of AA*). However, we show that the inequality is independent of the usual norm axioms. Some more general inequalities are also obtained.

NORMS AND CONDITION NUMBERS

By

Albert W. Marshall and Ingram Olkin

Boeing Scientific Research Laboratories and Stanford University

1. Summary and Introduction.

The genesis of this study is the proposition that under certain conditions, the matrix AA^* is more "ill-conditioned" than A. More precisely, the condition number $c_\varphi(A)$ is defined for non-singular matrices A as

$$c_\varphi(A) = \varphi(A)\, \varphi(A^{-1}) \, ,$$

where ordinarily φ is a norm. The statement concerning ill-conditioning of AA^* is the inequality

(c) $$c_\varphi(A) \leq c_\varphi(AA^*) \, .$$

Where $\varphi(A)$ is the maximum absolute characteristic root of A and where $\varphi(A) = (\operatorname{tr} AA^*)^{1/2}$, inequality (c) was proved by O. Tausky-Todd [6]. This raises the question of whether (c) is true for all norms. In this paper, we show that quite the contrary is true; (c) is independent of the usual norm axioms. However, we also prove that (c) does hold for a quite general class of norms.

In the course of proving these results, we obtain some inequalities for symmetric gauge functions, which may be of independent interest.

2. Gauge functions and matrix norms.

We call φ a matrix norm if

(aI) $\qquad\qquad \varphi(A) > 0 \qquad$ when $A \neq 0$,

(aII) $\qquad\qquad \varphi(\alpha A) = |\alpha| \, \varphi(A)$ for complex α ,

(aIII) $\qquad\qquad \varphi(A+B) \leq \varphi(A) + \varphi(B)$.

In addition to these basic axioms, various other conditions are sometimes imposed:

(aIV) $\qquad\qquad \varphi(E_{ij}) = 1$,

where E_{ij} is the matrix with one in the (i,j)-th position and zero elsewhere,

(aV) $\qquad\qquad \varphi(AB) \leq \varphi(A) \, \varphi(B)$,

(aVI) $\qquad \varphi(A) = \varphi(UA) = \varphi(AU)$ for all unitary matrices U.

If φ satisfies aI, aII, aIII, and aVI, φ is called a unitarily invariant norm.

There is an important connection between unitarily invariant norms and symmetric gauge functions. A function Φ on a complex vector space is called a gauge function if

(bI) $\qquad \Phi(u) > 0$ when $u \neq 0$,

(bII) $\qquad \Phi(\alpha u) = |\alpha| \; \Phi(u)$ for complex α ,

(bIII) $\qquad \Phi(u+v) \leq \Phi(u) + \Phi(v)$.

Often it is convenient to assume, in addition, that

(bIV) $\qquad \Phi(e_i) = 1$,

where e_i is the vector with one in the i-th place and zero elsewhere. If, in addition to bI, bII, and bIII,

(bV) $\qquad \Phi(u_1,\ldots,u_n) = \Phi(\epsilon_1 u_{i_1}, \; \ldots, \; \epsilon_n u_{i_n})$

whenever $\epsilon_j = \pm 1$ and $(i_1,\ldots,i_n)$ is a permutation of $(1,\ldots,n)$, then Φ is called a symmetric gauge function.

It was noted by Von Neumann [7] that a norm φ is unitarily invariant if and only if there exists a symmetric gauge function Φ such that $\varphi(A) = \Phi(\alpha)$ for all A, where $\alpha_1^2,\ldots,\alpha_n^2$ are the eigenvalues of AA*.

If Φ is a symmetric gauge function and u,v satisfy $u_i \leq v_i$, $i = 1,\ldots,n,$ then it follows [5, p. 85] that

(2.1) $\qquad \Phi(u_1,\ldots,u_n) \leq \Phi(v_1,\ldots,v_n)$.

If Φ is a symmetric gauge function satisfying bIV, then [5, p.86]

$$(2.2) \qquad \max_i |u_i| \leq \Phi(u_1,\ldots,u_n) \leq \sum_{i=1}^n |u_i| \, .$$

If φ is the unitarily invariant matrix norm determined by Φ as above, then it follows that

$$\frac{\varphi(AB)}{\varphi(A)\, \varphi(B)} \leq \frac{\sum_{i=1}^n \lambda_i (ABB^*A^*)}{[\max_i \lambda_i (AA^*)]\, [\max_j \lambda_j (BB^*)]} \leq n \, ,$$

where $\lambda_i(M)$ are the eigenvalues of M. Thus, for any $k \geq n$, $k\varphi$ is a unitarily invariant matrix norm also satisfying aV. Of course, φ itself satisfies aIV (since Φ satisfies bIV), and this property is destroyed by the renormalization.

3. The condition number inequality.

__Theorem 3.1.__ If φ is a unitarily invariant norm, then

$$(c) \qquad\qquad c_\varphi(A) \leq c_\varphi(AA^*) \, .$$

If Φ is a symmetric gauge function which determines φ, then we may rewrite (c) in the form

$$\Phi(\alpha_1,\ldots,\alpha_n)\, \Phi(\alpha_1^{-1},\ldots,\alpha_n^{-1}) \leq \Phi(\alpha_1^2,\ldots,\alpha_n^2)\, \Phi(\alpha_1^{-2},\ldots,\alpha_n^{-2}) \, .$$

Thus, Theorem 3.1 is a very special case of

<u>Theorem 3.2</u>. If Φ is a symmetric gauge function, then

$\Phi(\alpha_1^r,\ldots,\alpha_n^r)\,\Phi(\alpha_1^{-r},\ldots,\alpha_n^{-r})$ is increasing in $r > 0$, where $\alpha_i > 0$.

The proof of Theorem 3.2 is embodied in the lemmas below.

We say $(a_1,\ldots,a_n)$ is <u>majorized</u> by $(b_1,\ldots,b_n)$, written

(a) $\prec$ (b), if (i) $a_1 \geq \cdots \geq a_n > 0$, $b_1 \geq \cdots \geq b_n > 0$,

(ii) $\displaystyle\sum_1^k a_i \leq \sum_1^k b_i$, $k = 1,\ldots,n-1$, (iii) $\displaystyle\sum_1^n a_i = \sum_1^n b_i$.

<u>Lemma 3.3</u>. If (a) $\prec$ (b), and Φ is a symmetric gauge function, then

(3.1) $$\Phi(a_1,\ldots,a_n) \leq \Phi(b_1,\ldots,b_n) ,$$

(3.2) $$\Phi(a_1^{-1},\ldots,a_n^{-1}) \leq \Phi(b_1^{-1},\ldots,b_n^{-1}) .$$

<u>Proof</u>. A proof of (3.1) has been given by Fan [1]; by an argument similar to his, we prove (3.2).

First, note that we can assume for h and j fixed, $h < j$,

(3.3) $\quad a_h = \alpha b_h + (1-\alpha)b_j$, $a_j = (1-\alpha)b_h + \alpha b_j$, $a_i = b_i$, $i \neq h,j$.

That this is true follows from the fact that if (a) $\prec$ (b), then a can be derived from b by successive applications of a finite number of transformations of the form (3.3) (see [2, p. 47]).

Let $\widetilde{b} = (b_1, \ldots, b_h, b_j, b_{h+1}, \ldots, b_{j-1}, b_h, b_{j+1}, \ldots, b_n)$. By convexity,

$$(\alpha b_i + (1-\alpha)\widetilde{b}_i)^{-1} \leq \alpha b_i^{-1} + (1-\alpha)\widetilde{b}_i^{-1} .$$

Then using (2.1), it follows that

$$\Phi(a_1^{-1}, \ldots, a_n^{-1}) = \Phi(\alpha b_1 + (1-\alpha)\widetilde{b}_1)^{-1}, \quad \cdots \quad , (\alpha b_n + (1-\alpha)\widetilde{b}_n)^{-1}$$

$$\leq \Phi(\alpha b_1^{-1} + (1-\alpha)\widetilde{b}_1^{-1}, \quad \cdots \quad , \alpha b_n^{-1} + (1-\alpha)\widetilde{b}_n^{-1})$$

$$\leq \alpha \Phi(b_1^{-1}, \ldots, b_n^{-1}) + (1-\alpha) \Phi(\widetilde{b}_1^{-1}, \ldots, \widetilde{b}_n^{-1}) . \quad \|$$

As a consequence of Lemma 3.3., we have that if $(a) \prec (b)$ then

$$\Phi(a_1, \ldots, a_n) \Phi(a_1^{-1}, \ldots, a_n^{-1}) \leq \Phi(b_1, \ldots, b_n) \Phi(b_1^{-1}, \ldots, b_n^{-1}) .$$

The proof of Theorem 3.2 is completed by the following

<u>Lemma 3.4.</u> If $\alpha_1 \geq \ldots \geq \alpha_n > 0$ and $a_i = \alpha_i^r / \Sigma \, \alpha_j^r$, $b_i = \alpha_i^s / \Sigma \, \alpha_j^s$, $0 < r < s$, then $(a) \prec (b)$.

<u>Proof.</u> We must show that for all k,

$$\frac{\sum\limits_1^k \alpha_i^{\,r}}{\sum\limits_1^n \alpha_i^{\,r}} \leq \frac{\sum\limits_1^k \alpha_i^{\,s}}{\sum\limits_1^n \alpha_i^{\,s}} \quad , \qquad r < s \ ,$$

which is true if and only if

$$\sum_1^k \alpha_i^s \sum_{k+1}^n \alpha_j^r - \sum_1^k \alpha_i^r \sum_{k+1}^n \alpha_j^s = \sum_{i=1}^k \alpha_i^r \sum_{j=k+1}^n \alpha_j^r (\alpha_i^{s-r} - \alpha_j^{s-r}) \geq 0 .$$

The latter follows from $\alpha_i \geq \alpha_j$, $i < j$. $\|$

Observe that by (3.1) and Lemma 3.4, we have

$$\frac{\Phi(\alpha_1^r, \ldots, \alpha_n^r)}{\Phi(\alpha_1^s, \ldots, \alpha_n^s)} \leq \frac{\Sigma \alpha_i^r}{\Sigma \alpha_i^s} .$$

In view of (2.2), it is perhaps natural to expect that

$$(3.4) \quad \frac{\alpha_1^r}{\alpha_1^s} \leq \frac{\Phi(\alpha_1^r, \ldots, \alpha_n^r)}{\Phi(\alpha_1^s, \ldots, \alpha_n^s)} \leq \frac{\Sigma \alpha_i^r}{\Sigma \alpha_i^s} , \quad 0 < r < s, \quad \alpha_1 \geq \ldots \geq \alpha_n > 0 ,$$

for any symmetric gauge function Φ. To see this we need only prove
the left hand inequality, which may be written in the form

$$(3.5) \qquad \Phi([\frac{\alpha_1}{\alpha_1}]^s, \ldots, [\frac{\alpha_n}{\alpha_1}]^s) \leq \Phi([\frac{\alpha_1}{\alpha_1}]^r, \ldots, [\frac{\alpha_n}{\alpha_1}]^r) ,$$

and which is a consequence of (2.1).

An interesting counterpart to Theorem 3.2 can be obtained from (3.4).

Theorem 3.5. If Φ is a symmetric gauge function satisfying bIV, then

$[\Phi(\alpha_1^r, \ldots, \alpha_n^r)]^{\frac{1}{r}}$ is decreasing in $r > 0$ whenever $\alpha_i > 0$, $i = 1, 2, \ldots, n$.

8

Thus $[\Phi(\alpha_1^r,\ldots,\alpha_n^r)\,\Phi(\alpha_1^{-r},\ldots,\alpha_n^{-r})]^{\frac{1}{r}}$ is decreasing in $r > 0$.

Proof. We have that

$$1 \;\le\; \Phi([\frac{\alpha_1}{\alpha_1}]^s,\ldots,[\frac{\alpha_n}{\alpha_1}]^s) \;\le\; \Phi([\frac{\alpha_1}{\alpha_1}]^r,\ldots,[\frac{\alpha_n}{\alpha_1}]^r) \; ,$$

the first inequality by bIV and (2.1). The second inequality is (3.5).
Thus

$$\{\Phi([\frac{\alpha_1}{\alpha_1}]^s,\ldots,[\frac{\alpha_n}{\alpha_1}]^s)\}^r \;\le\; \{\Phi([\frac{\alpha_1}{\alpha_1}]^r,\ldots,[\frac{\alpha_n}{\alpha_1}]^r)\}^r \;\le\; \{\Phi([\frac{\alpha_1}{\alpha_1}]^r,\ldots,[\frac{\alpha_n}{\alpha_1}]^r)\}^s \; ,$$

so that

$$\{\Phi([\frac{\alpha_1}{\alpha_1}]^s,\ldots,[\frac{\alpha_n}{\alpha_1}]^s)\}^{\frac{1}{s}} \;\le\; \{\Phi([\frac{\alpha_1}{\alpha_1}]^r,\ldots,[\frac{\alpha_n}{\alpha_1}]^r)\}^{\frac{1}{r}} \; .$$

The theorem now follows from bII. ‖

Theorem 3.5 can, of course, be specialized to yield a kind of
converse to (c).

Theorem 3.6. If φ is a unitarily invariant norm satisfying aIV, then

(c*) $$[c_\varphi(AA^*)]^{1/2} \;\le\; c_\varphi(A) \; .$$

Condition (c*) can also be obtained under somewhat different
hypotheses. In particular, if φ satisfies aV, then

$$c_{\varphi}(AA^*) = \varphi(AA^*) \; \varphi((AA^*)^{-1}) \; \leq \; \varphi(A) \; \varphi(A^{-1}) \; \varphi(A^*) \; \varphi(A^{*-1}) = c(A) \; c(A^*) \; .$$

If also $\varphi(A) = \varphi(A^*)$, then (c*) follows. Of course, $\varphi(A) = \varphi(A^*)$ if φ is unitarily invariant.

4. Independence of the norm axioms and (c).

It is our purpose here to show that the condition number inequality (c) does not follow from the usual norm axioms aI - aV. In fact, aII, aIII, aIV, aV and (c) are independent.

Remark. It has been shown by Ostrowski [3] that aI is implied by aII, aIII, aV, together with $\varphi(A) \not\equiv 0$, so that aI is not included in the list of independent properties. Rella [4] has shown that aII, aIII, aIV and aV are independent, and we add (c) to this list.

The results which prove the independence of aII - aV and (c) are summarized in the following table, where + (-) indicates that a property is true (false).

$\varphi(A)$	aII	aIII	aIV	aV	(c)
1	-	+	+	+	+
$(\text{rank } A) \; (\text{tr } AA^*)^{1/2}$	+	-	+	+	+
$n \max |a_{ij}|$	+	+	-	+	+
$\max |a_{ij}|$	+	+	+	-	+
$\Sigma \, |a_{ij}|$	+	+	+	+	-

The remainder of this paper is devoted to proving the propositions indicated in the table.

The results for $\varphi(A) \equiv 1$ are obvious, so we begin by considering $\varphi(A) = (\text{rank } A)(\text{tr } AA^*)^{1/2}$. In this case, aII and aIV are obvious, and (c) follows from Theorem 3.1, since $(\text{tr } AA^*)^{1/2}$ is unitarily invariant. As is well known, $(\text{tr} AA^*)^{1/2}$ satisfies aV; this together with rank $AB \leq (\text{rank } A)(\text{rank } B)$ yields aV for $\varphi(A) = (\text{rank } A)(\text{tr } AA^*)^{1/2}$. That aIII is violated may be seen by taking $A = I$ and B the matrix with a unit in the $(1,1)$-th place and zeros elsewhere.

For $\varphi(A) = n \max\limits_{i,j} |a_{ij}|$ and $\max\limits_{i,j} |a_{ij}|$ the first four columns of the table are well known, and we need only prove (c). Let e_i be the row vector with one in the i-th position and zero elsewhere. Denote $M^{-1} = (m^{ij})$ where $M = (m_{ij})$, and let $U = AA^*$. By Cauchy's inequality,

$$|a_{ij}| \, |a^{\alpha\beta}| = |e_i A e_j^*| \, |e_\alpha A^{-1} e_\beta^*| \leq [(e_i U e_i^*)(e_j e_j^*)(e_\alpha e_\alpha^*)(e_\beta U^{-1} e_\beta^*)]^{1/2}$$
$$= (u_{ii} u^{\beta\beta})^{1/2} .$$

Hence,

$$\max\limits_{i,j} |a_{ij}| \max\limits_{\alpha,\beta} |a^{\alpha\beta}| \leq (\max\limits_{i} |u_{ii}| \max\limits_{\alpha} |u^{\alpha\alpha}|)^{1/2} ,$$

or

$$c_\varphi(A) \leq [c_\varphi(AA^*)]^{1/2} .$$

Since $U = AA^*$ is positive semi-definite,

$$u_{11}u^{11} = (e_1 U e_1^*)(e_1 U^{-1} e_1^*) \geq (e_1 e_1^*)^2 = 1 ,$$

and it follows that $c_\varphi(AA^*) \geq 1$. Thus, we have that

$$(4.1) \qquad c_\varphi(A) \leq [c_\varphi(AA^*)]^{1/2} \leq c_\varphi(AA^*) ,$$

which gives (c).

Note that the left inequality of (4.1) is a reversal of inequality (c*). That (4.1) also holds if $\varphi(A)$ is the maximum of the absolute values of the characteristic values of A was proved by O. Tausky-Todd [6].

Since the first four columns of the table are well known for

$\varphi(A) = \Sigma|a_{ij}|$, we again need consider only (c). If $A = \begin{pmatrix} B & 0 \\ 0 & n^{-1} I \end{pmatrix}$,

where $B = \begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix}$. Then (c) is violated.

BIBLIOGRAPHY

[1] Fan, K. (1951). Maximum Properties and Inequalities for the
     Eigenvalues of Completely Continuous Operators,
     <u>Proc. Nat. Acad. Sci. U.S.A.</u> <u>37</u>, 760-766.

[2] Hardy, G. H., Littlewood, J. E., and Pólya, G., (1959) <u>Inequalities</u>
     (2nd. Ed.), Cambridge University Press, Cambridge.

[3] Ostrowski, A. (1955). Über Normen Von Matrizen, <u>Math. Z.</u>, <u>63</u>
     2-18.

[4] Rella, T. (1936). Über den Absoluten Betrig Von Matrizen,
     <u>Proc. of the International Congress of Mathematicians,</u>
     Oslo, 29-31.

[5] Schatten, R. (1950) <u>A Theory of Cross-spaces,</u> Princeton
     University Press, Princeton.

[6] Tausky-Todd, O., (1950) Note on the Condition of Matrices,
     <u>Math. Tables Aids Comput.</u>, <u>V. 4</u>, p. 111-112.

[7] Von Neumann, (1937) Some Matrix-Inequalities and Metrization
     of Matric-Space, <u>Tomsk. Univ., Rev. Vol. 1</u>, p. 286-300,
     (in Collected Works, Vol. IV, Pergamon Press, 1962).